

Sultan Qaboos University

College of Science

Department of Statistics

STAT5521: Categorical data Analysis Fa11 21

**Studing some effect Of Low Birthweight**

**Done by:**

Asma Said Almoqbali (ID:124033)

**Introduction**

Low birth weight is an outcome that has been of concern to physicians for years. This is due to the fact that infant mortality rates and birth defect rates are very high for low birth weight babies. A woman's behavior during pregnancy can greatly affect baby of normal birth weight. My primary objective was to find meaningful relationships between a given mother’s health state and the birth weight of her child, and develop a model that could reasonably predict said birth weight. Previous studies have shown that variables such as smoking status and prior premature labors can have a significant impact on a newborn’s overall health and weight, and I wanted to confirm these correlations for myself so that I may better educate and consult future parents about maintaining the health of their unborn children throughout the pregnancy process.

the data was collected were collected at Baystate Medical Center, Springfield MA, during 1986

"https://raw.githubusrcontent.com/CEHS-research/data/master/Regression/lowbwt.txt”

**Problem Statement**

Low birth weight can cause serious health problems for some babies. A baby with a low birth weight may have trouble eating, gaining weight and fighting infections. Some babies with low birth weight may have long-term health problems, too. The aim of the study was to explore the extent of low birth weight. One of the research goals is to find out whether low birth weight has an effect on his parents. Moreover, studying the causes of low birth weight helps in raising awareness among the parents and thus reducing the number of birth weights below the normal birth weight

We analyzed the Low birth weight data of whether each of n = 189 childe has Birth Weight => 2500g with (y=0) or Birth Weight < 2500g with(y=1)

Dependent variable or outcome

Lbw Low birth weight (outcome (binary))

0 = birth weight >2500 g (normal)

1 = birth weight < 2500 g (low))

Independent variables or predictors

Age age of mother, in years (numerical)

Smoke Smoking status during pregnancy:1 = Yes, 0 = No (binary)

ptl History of premature labor: 0 = None, 1 = One, 2 = two, 3 = three

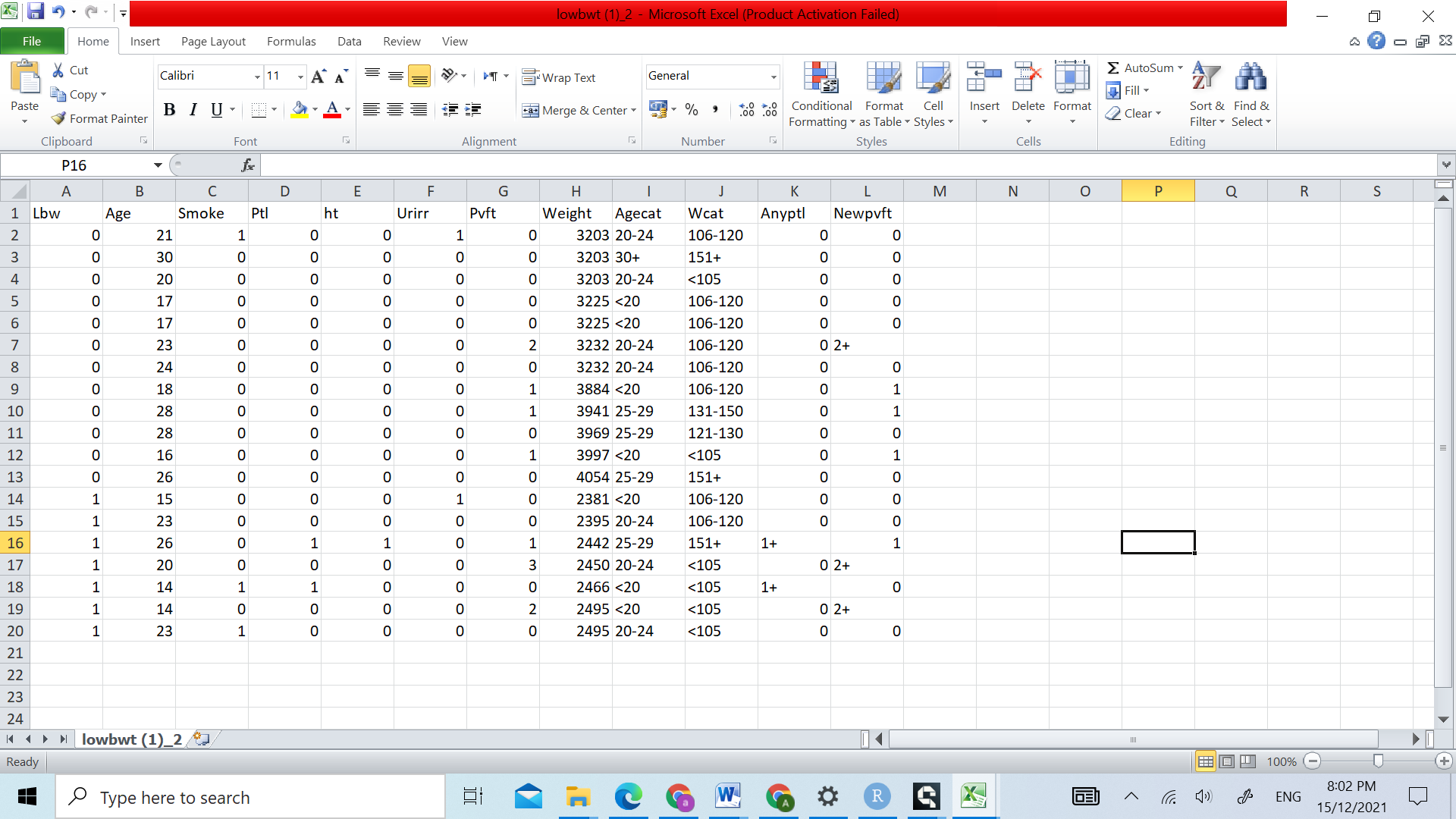
ht History of hypertension: 1 = Yes, 0 = No(binary)

I choose from The low Birth weight data file four explanatory variables( that I thought there effect is high): age of the mother as a numerical variable, if mother Smoking Status During Pregnancy (1 = Yes, 0 = No) , History of Premature Labor (0 = None, 1 = One, etc.) but we will disgust it to two category only (0=none and 1=at least one ) , History of Hypertension (1 = Yes, 0 = No). We now explore logistic regression modeling using all of them. We first fit a model that contains all the main effects, treating all of them as qualitative (nominal-scale) factors. Let (s1, s2)be indicator variables for smoke and let (p1,p2) be indicator variables for premature and (h1,h2)for hypertension.

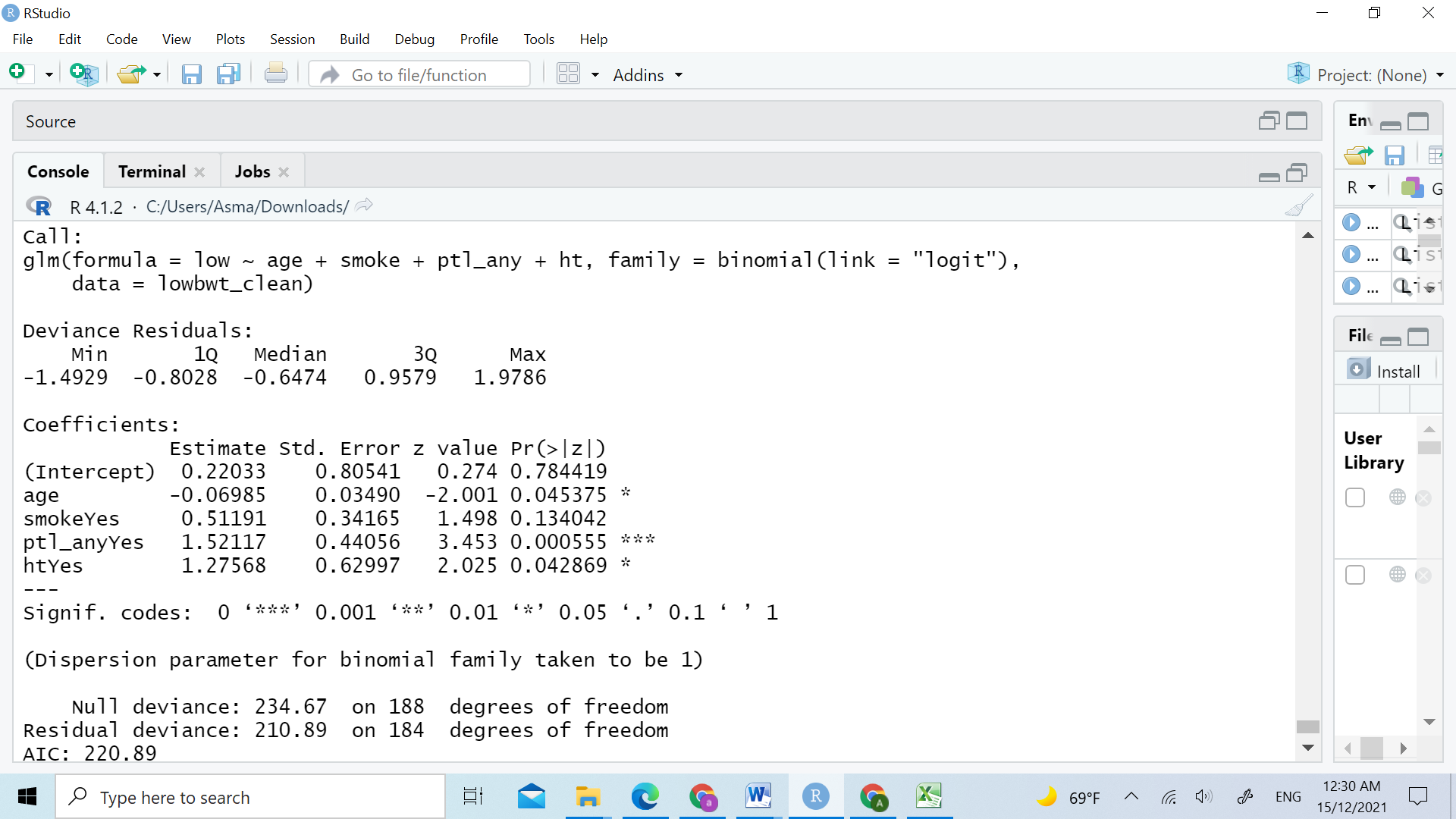
The model is (The most popular model for binary data is logistic regression)

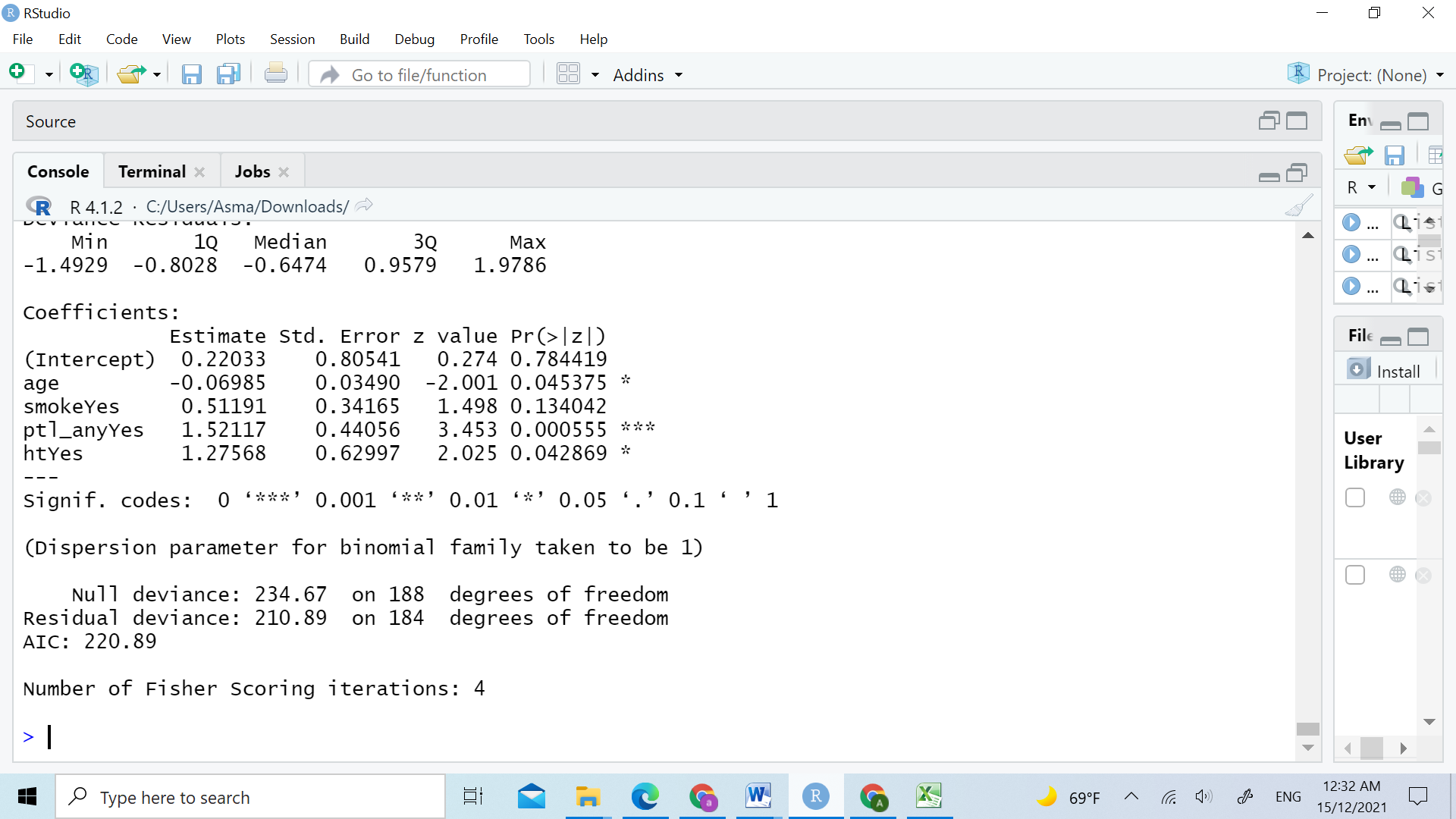
logit[P(Y = 1)] = α + β1(age)+ β2s1+ β3s2 + β4p1 + β5p2+ β6h1+ β7h2.

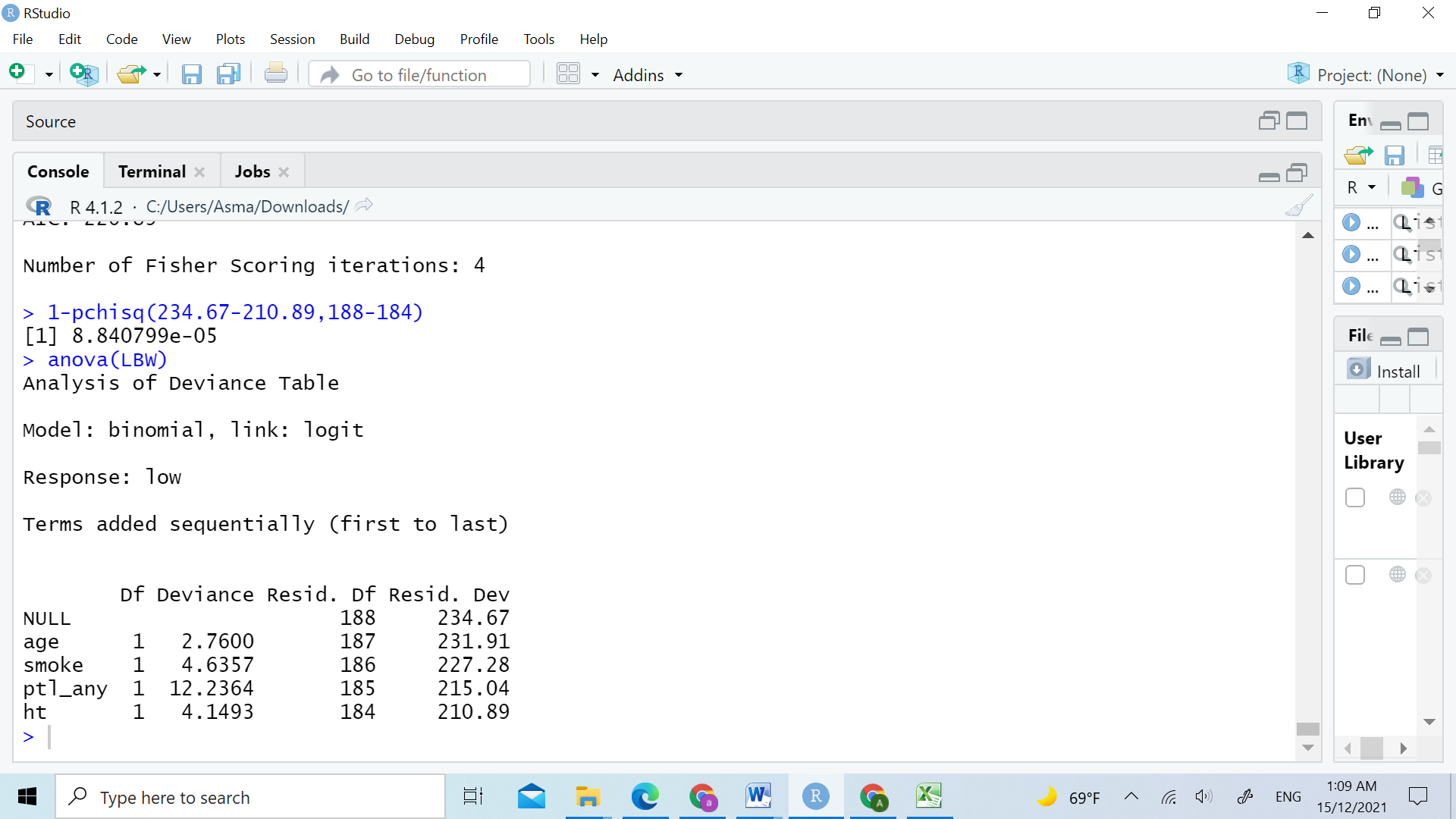
Part of data:



Here are some results, using R







statistical inference for the logistic regression model parameters helps us judge the significance and size of the effects of explanatory variables. We use the Wald and likelihood-ratio A likelihood-ratio test that Y is jointly independent of the four explanatory variables simultaneously tests H0: β1 = ··· = β4 = 0. The test statistic is the difference between the null deviance and the residual deviance, which is 234.67-210.89= 23.78 with df = 188-184=4

This shows evidence that at least one explanatory variable has an effect (P-value = 0 < 0.05). Although this overall test is highly significant, the test results for individual variables are discouraging. Likelihood-ratio tests for individual explanatory variables show only one variable is not significant at the 0.05 level (smokeyes). The P-value for the overall test is small, yet the lack of significance for individual effects is a warning sign of multicollinearity.

Significant testing:

For age, H0: β1 = 0 vs Ha: β1 ≠ 0 p-value = 0.045

for smokeyes, H0: β2 = 0 vs Ha: β2 ≠ 0 p-value = 0.134

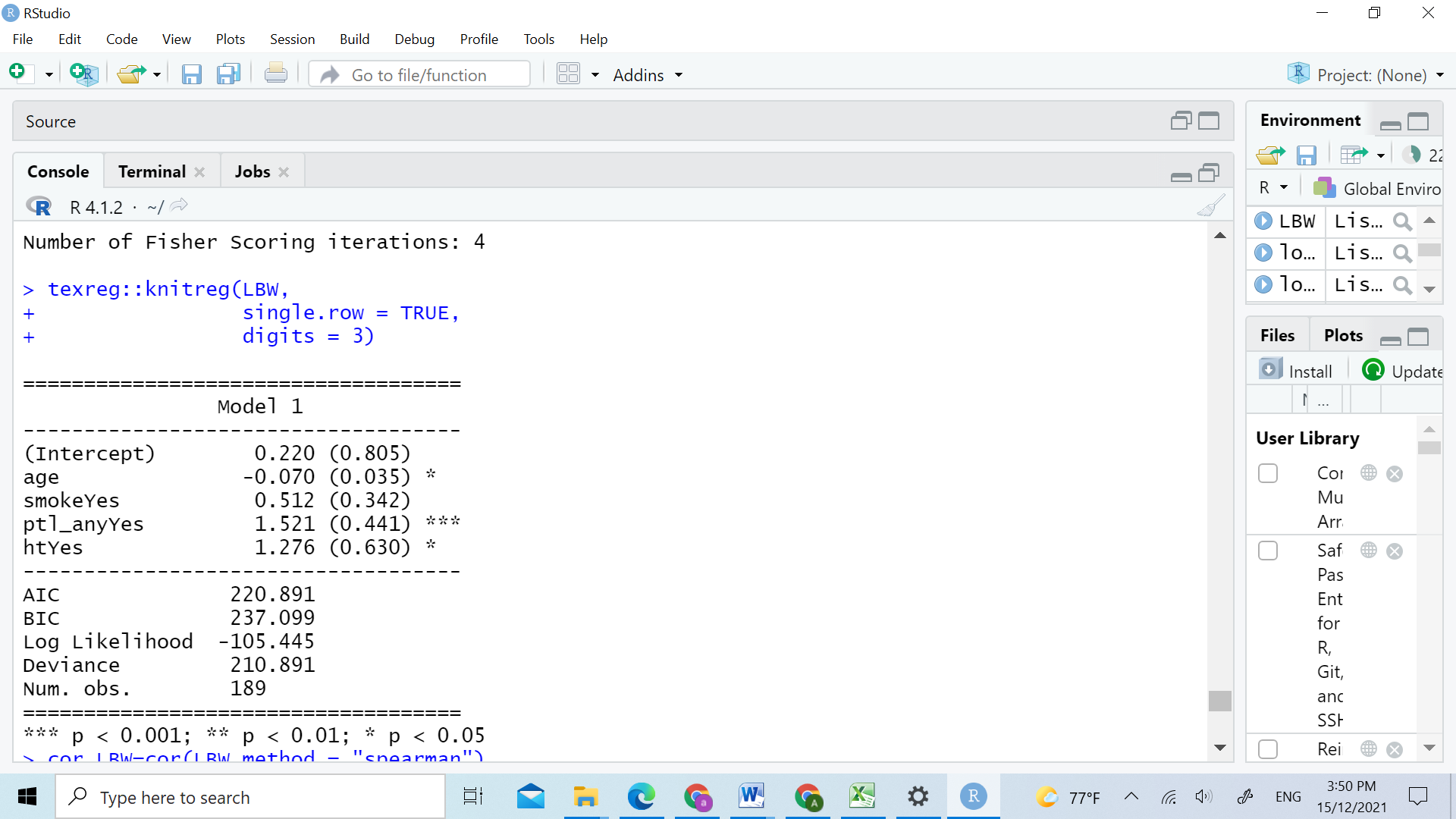
For ptl\_any, H0: β3 = 0 vs Ha: β3 ≠ 0 p-value = 0.000555

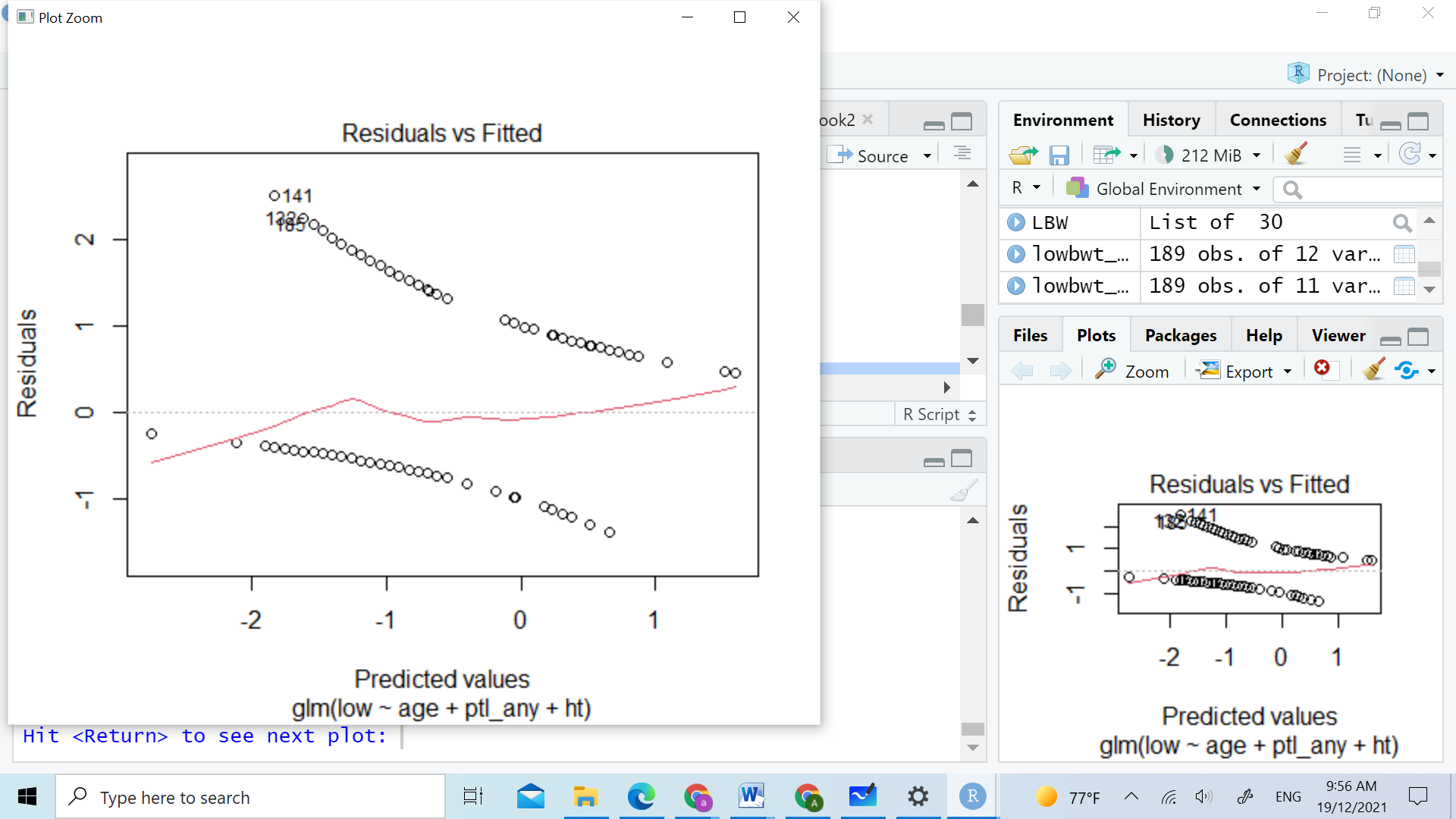
for htyes, H0: β4 = 0 vs Ha: β4 ≠ 0 p-value = 0.042

Shows weak evidence of low Birth weight (P-value > 0.05) and strong evidence of other variables effect (P-value < 0.05). Our further analysis drops low birth weight and uses other variables as the potential explanatory variables. - The estimated odd of having low birth weight increases as the smoke group move from one category (no) to another (yes). - The Wald statistic z = 1.498 shows strong evidence of positive effect of smoke group on the low birth weight (p-value < 0.0001).

And The Wald statistic z = 3.453 shows strong evidence of positive effect of the History of premature labor on the birth weight (p-value < 0.0001).

And at the code below we show clearly parameters are in terms of the ‘logit’ or log odds ratio for each variable





(from the plot the model is not linear well)

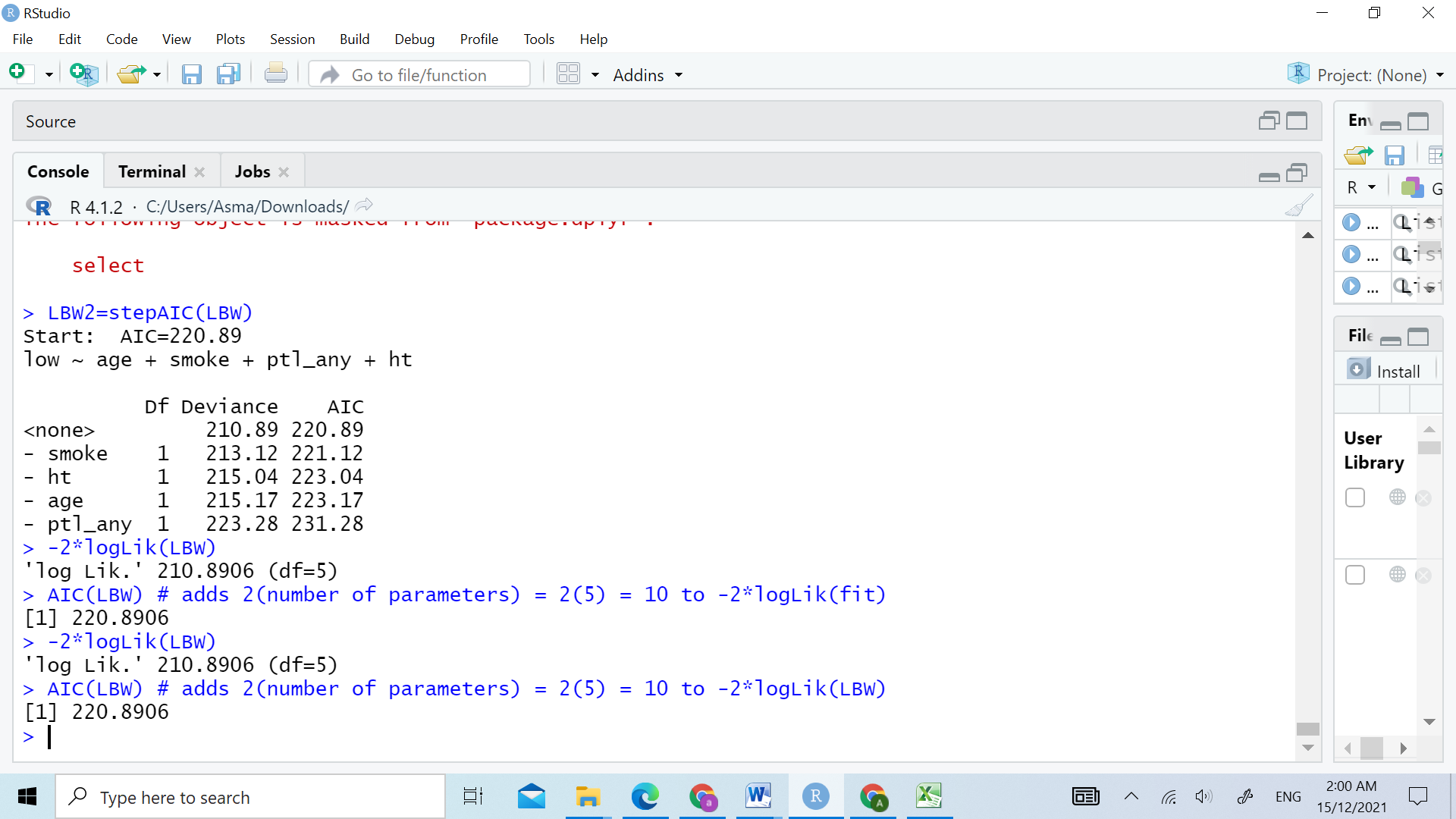
**AIC and the Bias/Variance Tradeoff**

The best known is the Akaike information criterion (AIC). It judges a model by how close its fitted values tend to be to the true expected values, as summarized by a certain expected distance between the two. The optimal model, which tends to have its fitted values closest to the true outcome probabilities, is the one that minimizes

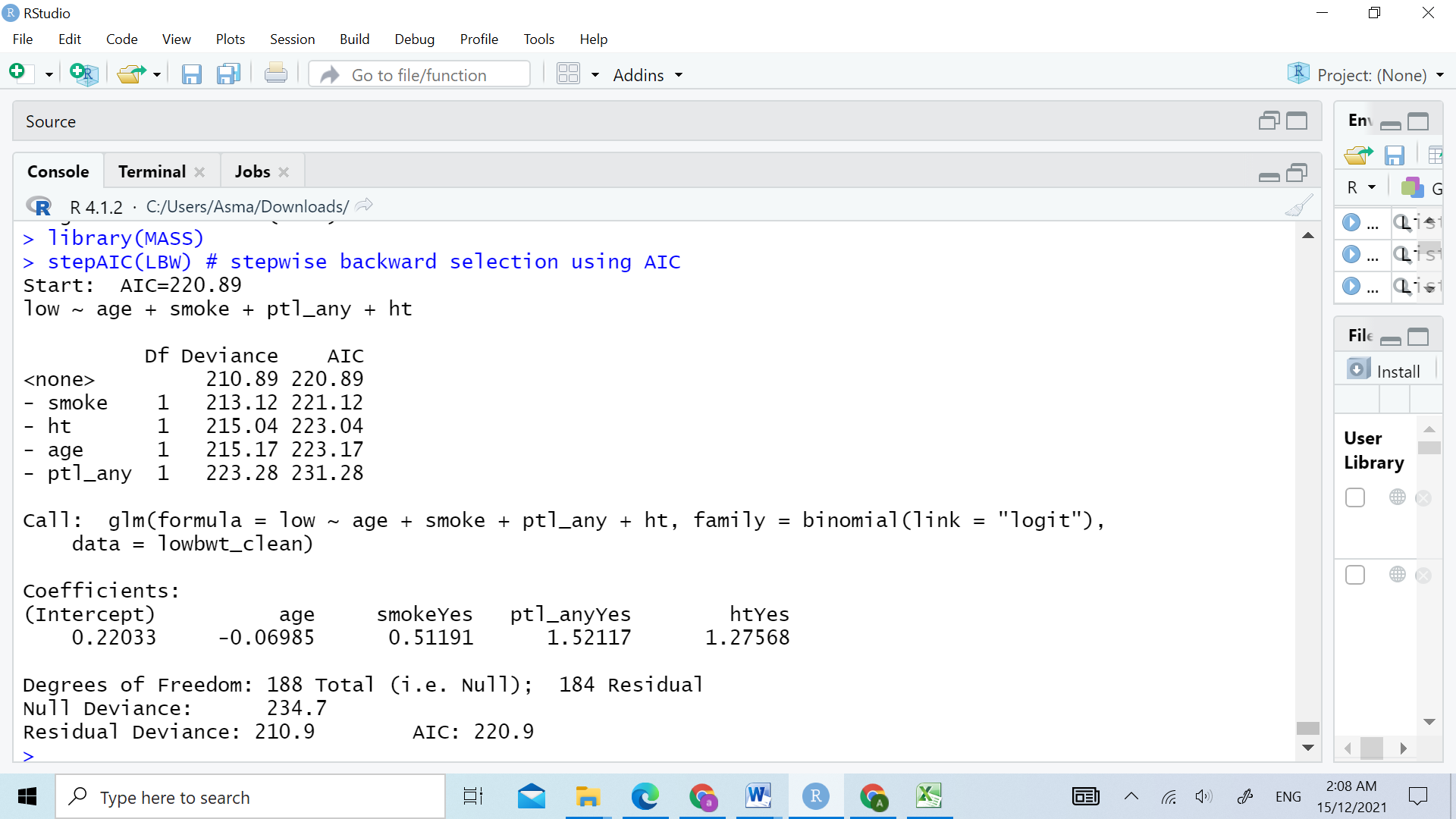
AIC = −2(log likelihood)+2(number of parameters in model).

Because smaller AIC is better, the AIC penalizes a model for having many parameters. Even though a simpler model necessarily has smaller log likelihood (and larger deviance) than a more complex model, the simpler model may well provide better estimates of the true expected values

Using R



The model has 5 parameters — an intercept with all 4 explanatory variables. Thus, AIC = 210.8906+ 2(5) = 220.8906. Of models, AIC is smallest for that model. AIC can also be the basis of stepwise model selection. For instance, the next R output uses AIC to select a model in a backward manner. We start with all four potential explanatory variables as main effects. At each step we remove the variable so that AIC decreases the most, until we get to the stage in which AIC increases if we remove any other variables:



By using AIC to identify which model is the best, we have found that the this model which contains age, smoke, History of premature labor, History of hypertension as expletory variables is the best model since it has the lowest AIC =220.9 . This model shows that the effect of age, smoke, History of premature labor, History of hypertension is highly significant on low birth weight

**Conclusion:**

As we can see from the study results and analysis that the birth weight directly affected by mother . Because of the tremendous advances in care of sick and premature babies, more and more babies are surviving despite being born early and being born very small. However, prevention of preterm births is one of the best ways to prevent babies born with low birth weight. Prenatal care is a key factor in preventing preterm births and low birth weight babies. At prenatal visits, the health of both mother and fetus can be checked. Because maternal nutrition and weight gain are linked with fetal weight gain and birth weight. Mothers should also avoid alcohol, cigarettes, and illicit drugs, which can contribute to poor fetal growth, among other complications.